

EXAMEN  
VIERNES 15 DE JULIO DE 2016.

Número de Examen	Cédula	Nombre y Apellido

PARA USO DOCENTE.			
Ej. 1	Ej. 2	Ej. 3	TOTAL

**Ejercicio 1. [30 puntos]**

Se estima que en Uruguay hay 40.000 personas celíacas. Se asume que la población del Uruguay es de 3.400.000 habitantes.

- Para una muestra aleatoria de 500 personas:
  - ¿Cuál es la probabilidad de que no haya ninguna persona celíaca en la muestra?
  - ¿Y de qué haya exactamente una persona celíaca?
  - ¿Cuántas personas celíacas esperaría encontrar en esta muestra?
- Un primer indicio de la enfermedad celíaca es la presencia de un número elevado de anticuerpos IgA endomysial en una muestra de sangre. Se sabe que para dicho examen de sangre:
  - la probabilidad de que el examen sea positivo dado que la persona es celíaca es de 0.92 (se denomina sensibilidad del examen).
  - la probabilidad de que el examen sea negativo dado que la persona no es celíaca es de 0.985 (se denomina especificidad del examen)

Calcular las siguientes probabilidades:

- Probabilidad de que el examen sea positivo dado que la persona no es celíaca.
  - Probabilidad de que el examen sea negativo dado que la persona es celíaca.
  - Probabilidad de que el examen sea positivo. Idem para examen negativo.
  - Probabilidad de que la persona sea celíaca dado que el examen es positivo.
- Otra posibilidad para detectar la enfermedad celíaca en una muestra de sangre es la presencia de anticuerpos IgA tissue transglutaminase. En la tabla siguiente se muestran los resultados obtenidos para una muestra aleatoria de mil personas:

	Examen Positivo	Examen Negativo
Celíaco	11	1
No Celíaco	40	948

Calcular la sensibilidad y especificidad de este examen.

**Ejercicio 1. Solución**

De los datos de personas celíacas en Uruguay obtenemos que

$$\mathbb{P}(\text{persona celíaca}) = \frac{40.000}{3.400.000} = \frac{4}{340} \simeq 0.012.$$

1. Dada una muestra aleatoria de 500 personas uruguayas, se define  $X$  como la cantidad de personas celíacas en la muestra. Puede considerarse que la variable aleatoria  $X$  tiene distribución Binomial con parámetros  $n = 500$  y  $p = 0.012$ . Por lo tanto,

- (a)  $\mathbb{P}(X = 0) = (0.988)^{500} \simeq 0.0024$ .
- (b)  $\mathbb{P}(X = 1) = 500 \times 0.012 \times (0.988)^{499} \simeq 0.0145$ .
- (c) Sabemos que si  $X \sim \text{Bin}(n, p)$  entonces  $\mathbf{E}(X) = np$ . En este caso se tiene entonces que la cantidad esperada de personas celíacas en la muestra es  $\mathbf{E}(X) = 500 \times 0.012 = 6$ .

2. Se definen los siguientes sucesos:

- $E^+ = \{\text{el examen es positivo}\}$  y  $E^- = \{\text{el examen es negativo}\}$
- $C = \{\text{la persona es celíaca}\}$  y  $NC = \{\text{la persona no es celíaca}\}$

De la letra, sabemos que  $\mathbb{P}(E^+ | C) = 0.92$  y  $\mathbb{P}(E^- | NC) = 0.985$ , de donde se obtiene que:

- (a)  $\mathbb{P}(E^+ | NC) = 1 - 0.985 = 0.015$
- (b)  $\mathbb{P}(E^- | C) = 1 - 0.92 = 0.08$
- (c)

$$\begin{aligned} \mathbb{P}(E^+) &= \mathbb{P}(E^+ | C) \times \mathbb{P}(C) + \mathbb{P}(E^+ | NC) \times \mathbb{P}(NC) \\ &= 0.92 \times 0.012 + 0.015 \times 0.988 = 0.0256. \end{aligned}$$

$$\mathbb{P}(E^-) = 1 - \mathbb{P}(E^+) = 1 - 0.0256 = 0.9744$$

(d)

$$\mathbb{P}(C | E^+) = \frac{\mathbb{P}(E^+ | C) \times \mathbb{P}(C)}{\mathbb{P}(E^+)} = \frac{0.92 \times 0.012}{0.0256} = 0.43125$$

3. Por definición se tiene que:

- Sensibilidad del examen =  $\mathbb{P}(E^+ | C) = \frac{11}{12} = 0.9167$ ,
- Especificidad del examen =  $\mathbb{P}(E^- | NC) = \frac{948}{988} = 0.9595$ .

## Ejercicio 2. [40 puntos]

Se desea realizar un estudio de las precipitaciones diarias (i.e. lluvias) para una obra vial en una determinada región. La variable de interés es la precipitación diaria máxima  $M$  en un determinado mes del año. Esto es, se registran las precipitaciones de cada día del mes (en mm), y se define  $M$  como el valor máximo registrado.

1. De experiencias anteriores se sabe que la variable aleatoria  $X = \log(M)$  (el logaritmo en base  $e$  de  $M$ ) tiene distribución normal  $N(\mu, \sigma^2)$ .

- (a) Hallar la mediana de  $M$  en función de  $\mu$ .
- (b) Probar que la esperanza de  $M$  es  $e^{\mu + \frac{\sigma^2}{2}}$ .

*Sugerencia:* usar la igualdad  $\frac{-x^2 + 2(\mu + \sigma^2)x - \mu^2}{2\sigma^2} = -\frac{(x - (\mu + \sigma^2))^2}{2\sigma^2} + \mu + \frac{\sigma^2}{2}$ .

- (c) Hallar la función de distribución y la densidad de  $M$ .

2. La Tabla 1 muestra la precipitación diaria máxima en la ciudad de Wageningen (Holanda) correspondiente al mes de Noviembre, durante los 19 años que van de 1948 a 1966.

- (a) A partir de los datos de la Tabla 1, realizar un diagrama de caja (boxplot) para  $M$  y un histograma para  $X$ . El histograma hacerlo con 7 intervalos que cubran el rango de 2.0 a 5.5. Indicar todos los valores numéricos necesarios para hacer ambos gráficos. Estudiar la simetría de ambas distribuciones.

$M_i$ (en mm)	10	28	30	36	38	40	45	46	56	57	65	65	67	74	92	99	111	158	200
$X_i = \log(M_i)$	2.30	3.33	3.40	3.58	3.64	3.69	3.81	3.83	4.03	4.04	4.17	4.17	4.20	4.30	4.52	4.60	4.71	5.06	5.30

Table 1: Precipitación diaria máxima (en mm) del mes de Noviembre durante los años 1948-1966. Los datos ya están ordenados de forma creciente. La segunda fila muestra los logaritmos de las precipitaciones. Datos adicionales que pueden ser de utilidad:  $\sum_{i=1}^{19} M_i = 1317$ ,  $\sum_{i=1}^{19} M_i^2 = 130615$ ,  $\sum_{i=1}^{19} X_i = 76.68$  y  $\sum_{i=1}^{19} X_i^2 = 317.74$ .

- (b) Hallar los estimadores de máxima verosimilitud de  $\mu$  y  $\sigma^2$ . Calcularlos para los valores de la Tabla 1.
- (c) Dar dos estimaciones diferentes para el valor esperado de  $M$  para el mes de Noviembre.

## Ejercicio 2. Solución

1. (a) Buscamos  $m_M \in \mathbb{R}$  tal que  $\mathbb{P}(M \leq m_M) = 1/2$ , esto es:

$$\mathbb{P}(M \leq m_M) = \mathbb{P}(\log(M) \leq \log(m_M)) = \mathbb{P}(X \leq \log(m_M)) = \frac{1}{2}$$

Sabemos que  $X$  tiene distribución  $N(\mu, \sigma^2)$ , entonces la mediana de  $X$  es el parámetro  $\mu$ , de donde se deduce que:

$$\log(m_M) = \mu \rightarrow m_M = e^\mu.$$

- (b) Si consideramos como  $g(x) = e^x$  tenemos que:

$$\begin{aligned} \mathbb{E}(M) &= \mathbb{E}(e^X) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^x e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{x + \frac{-x^2 + 2x\mu - \mu^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{x^2 + 2(\mu + \sigma^2)x - \mu^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x - (\mu + \sigma^2))^2}{2\sigma^2} + \mu + \frac{\sigma^2}{2}} dx = e^{\mu + \frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x - (\mu + \sigma^2))^2}{2\sigma^2}} dx = e^{\mu + \frac{\sigma^2}{2}}. \end{aligned}$$

- (c) Función de distribución de  $M$ :

$$F_M(x) = \mathbb{P}(M \leq x) = \mathbb{P}(\log(M) \leq \log(x)) = \mathbb{P}(X \leq \log(x)) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)$$

donde  $\Phi$  es la función de distribución de una v.a.  $\mathcal{N}(0, 1)$  La función de densidad de  $M$  la obtenemos derivando la función de distribución. Por lo cual,

$$f_M(x) = (F_M(x))' = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2} \frac{(\log(x) - \mu)^2}{\sigma^2}}.$$

2. (a) Ver figura 1. Se puede observar que  $M$  es claramente no simétrica mientras  $X$  por corresponder a datos normales, presenta mayor simetría. Es posible observar además que mientras la mediana y el promedio de los valores de  $X$  son casi idénticos, siendo estos valores muy diferentes para  $M$ .
- (b) La función de verosimilitud es:

$$\begin{aligned} L_M(\mu, \sigma^2) &= \prod_{i=1}^n f_M(M_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{M_i} e^{-\frac{1}{2} \frac{(\log(M_i) - \mu)^2}{\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \frac{1}{\prod_{i=1}^n M_i} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log(M_i) - \mu)^2} \end{aligned}$$

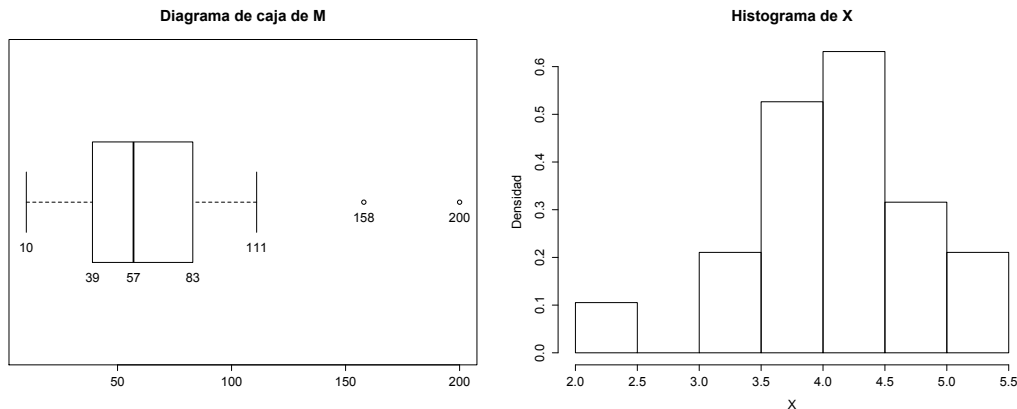


Figure 1: Bopxplot para  $M$  e histograma para  $X$ .

Tomando logaritmo resulta que:

$$g(\mu, \sigma^2) = \log L_M(\mu, \sigma^2) = -n \log(\sqrt{2\pi\sigma^2}) - \log\left(\prod_{i=1}^n M_i\right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log(M_i) - \mu)^2$$

Derivando, respecto de  $\mu$  y de  $\sigma^2$  resulta que:

$$\begin{cases} \partial_{\mu} g(\mu, \sigma^2) = \frac{1}{2\sigma^2} \sum_{i=1}^n (\log(M_i) - \mu) \\ \partial_{\sigma^2} g(\mu, \sigma^2) = -\frac{n}{\sigma^3} + \frac{1}{\sigma^3} \sum_{i=1}^n (\log(M_i) - \mu)^2 \end{cases}$$

Anulando las derivadas para hallar los extremos, se deduce que los estimadores de máxima verosimilitud son:

- $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(M_i)$
- $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (\log(M_i) - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (\log(M_i) - \frac{1}{n} \sum_{i=1}^n \log(M_i))^2$

Utilizando los datos de la Tabla 1, resulta que:

$$\begin{cases} \hat{\mu} = \frac{1}{19} 76.68 \cong 4.04 \\ \widehat{\sigma^2} = \frac{1}{19} (317.74 - 19 \cdot (\frac{76.68}{19})^2) \cong 0.436 \end{cases}$$

son los valores de estimadores de máxima verosimilitud de  $\mu$  y  $\sigma^2$  respectivamente.

(c) Dos maneras de estimar  $E(M)$  son:

- Ley de los grandes números para los datos  $M_i$ :  $\overline{M}_{19} = 69.31$
- Utilizar que  $g(x) = e^x$  es una función continua y por lo tanto usando la parte anterior resulta que  $e^{\hat{\mu} + \frac{\widehat{\sigma^2}}{2}} = e^{4.04 + \frac{0.436}{2}} = 70.67$  es también un estimador de  $E(M)$ .

### Ejercicio 3. [30 puntos]

Cuando el precio de las acciones de una compañía sube respecto al precio del día anterior, decimos que ha ocurrido un incremento en las acciones de dicha compañía. Denotamos por  $S$  la variable que cuenta el número de incrementos en un mes. Se sabe que la distribución de  $S$  se ajusta bien a una distribución de Poisson de parámetro  $\lambda$ . La Tabla 2 muestra los datos de la cantidad de incrementos del precio de las acciones de la compañía Google por mes, en los meses que van desde Enero del 2006 hasta Diciembre del 2010.

- A partir de los datos de la Tabla 2 estimar la esperanza y la varianza de  $S$ .
- En lo que sigue se asume que  $n = 60$  es un valor grande y es posible utilizar aproximaciones.

Nro. de incrementos en el mes	5	6	7	8	9	10	11	12	13	14	15	16
Frecuencia	1	6	2	9	9	10	5	8	5	4	0	1

Table 2: La tabla muestra la cantidad de incrementos del precio de las acciones de Google por mes, y la frecuencia con qué ocurrieron, durante los 60 meses que van de Enero 2006 a Diciembre 2010. Por ejemplo, de la tercer columna de la tabla deducimos que en 2 meses (del total de 60 meses estudiados) se observaron 7 incrementos en las acciones.

- (a) Hallar un intervalo de confianza al nivel 0.9 para  $\lambda$ .  
(b) Se considera el siguiente test de hipótesis

$$\begin{cases} H_0 : \lambda = 10 \\ H_1 : \lambda < 10 \end{cases}$$

- i. ¿Cuál es la decisión para el nivel  $\alpha = 0.1$ ?  
ii. Hallar el  $p$ -valor. ¿Cambia la respuesta si la hipótesis alternativa se cambia por  $H_1 : \lambda \leq 9.95$ ?  
iii. Para la hipótesis alternativa  $H_1 : \lambda = 9.95$  y  $\alpha = 0.1$ , hallar la probabilidad de error de tipo II.

### Ejercicio 3. Solución

1. Utilizando la Tabla 2, se tiene que  $p_n(i) = \frac{1}{60} \times \text{Frecuencia ocurrencia}(i)$  es una estimación de  $\mathbb{P}(S = i)$ . Por lo tanto  $\hat{\lambda} = \bar{S}_n = \sum_{i=5}^{16} ip_n(i) = 9.93$  es una estimación de la esperanza de  $S$ . Al tratarse de una distribución Poisson, sabemos que  $\text{Var}(S) = \mathbb{E}(S)$  por lo tanto 9.93 es también una estimación de  $\text{Var}(S)$ .

Otra forma de estimar la varianza es usando  $S_n^2$ , donde

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{60} S_i = \frac{n-1}{n} \sum_{i=5}^{16} (i - \mathbb{E}(S))^2 p_n(i) = 6.19.$$

2. (a) El intervalo de confianza aproximado para  $\lambda = \mathbb{E}(S)$  a nivel  $1 - \alpha$  es:

$$I_\alpha = \left[ \bar{S}_n - \frac{z_{\alpha/2} \hat{\sigma}}{\sqrt{n}}, \bar{S}_n + \frac{z_{\alpha/2} \hat{\sigma}}{\sqrt{n}} \right]$$

donde  $\hat{\sigma}^2$  es una estimación de la varianza. Para  $\alpha = 0.1$ , resulta que  $z_\alpha = 1.64$  y se obtienen los siguientes intervalos:

- Usando  $\hat{\sigma} = \sqrt{\hat{\lambda}} = 3.15$  se obtiene que  $I_\alpha = [9.26; 10.6]$
  - Usando  $\hat{\sigma} = S_n = 2.49$  se obtiene que  $I_\alpha = [9.4; 10.46]$
- (b) i. Usando que bajo  $H_0$  los datos se asumen Poisson de parámetro  $\lambda_0 = 10$ , se plantea la región crítica  $\mathcal{R}_\alpha = \{\bar{S}_n < \lambda_0 - \frac{z_\alpha \sqrt{\lambda_0}}{\sqrt{n}}\}$ . Utilizando que  $\lambda = 10$  y  $z_{0.1} = 1.28$ , resulta que  $\mathcal{R}_\alpha = \{\bar{S}_n < 9.47\}$ . Dado que  $\bar{S}_n = 9.93$ , se concluye que no es posible rechazar  $H_0$ .  
Como antes también es posible utilizar  $S_n$  como estimador de la varianza. En este caso, se obtiene la región crítica  $\mathcal{R}_\alpha = \{\bar{S}_n < \lambda_0 - \frac{z_\alpha S_n}{\sqrt{n}}\} = \{\bar{S}_n < 9.59\}$  y nuevamente no es posible rechazar  $H_0$ .
- ii. Por definición de  $p$ -valor tenemos que:

$$\begin{aligned} \alpha^* &= \mathbb{P}(\bar{S}_n \leq 9.93 | H_0 \text{ cierta}) = \mathbb{P}(\bar{S}_n \leq 9.93 | \lambda = 10) = \Phi\left(\frac{9.93 - 10}{\sqrt{10}} \sqrt{60}\right) \\ &= \Phi(-0.17) = 0.432. \end{aligned}$$

Una vez más es posible utilizar  $S_n$  en la normalización. En este caso se obtiene que

$$\alpha^* = \Phi\left(\frac{9.93 - 10}{\sqrt{6.19}}\sqrt{60}\right) = \Phi(-0.217) = 0.413.$$

La hipótesis alternativa no influye en el p-valor.

- (c) Para  $\alpha = 0.1$  se tiene que  $\mathcal{R}_\alpha = \{\bar{S}_{60} \leq 9.47\}$ . Por lo tanto la probabilidad de error de tipo II es:

$$\beta = \mathbb{P}_{H_1}(H_0) = \mathbb{P}_{H_1}(\bar{S}_{60} > 9.47 | \lambda = 9.95) = 1 - \Phi\left(\frac{9.47 - 9.95}{\sqrt{10}}\sqrt{60}\right) = 1 - \Phi(-1.13) = 0.87.$$

Utilizando  $S_n$  en la normalización se obtiene:  $\mathcal{R}_\alpha = \{\bar{S}_{60} \leq 9.59\}$ . Por lo tanto la probabilidad de error de tipo II es:

$$\beta = \mathbb{P}_{H_1}(H_0) = \mathbb{P}_{H_1}(\bar{S}_{60} > 9.59 | \lambda = 9.95) = 1 - \Phi\left(\frac{9.59 - 9.95}{\sqrt{6.199}}\sqrt{60}\right) = 1 - \Phi(-1.06) = 0.856.$$